OCUL Ontario Council of University Libraries

OCUL Artificial Intelligence/Machine Learning Report and Strategy

OCUL Task Force on Machine Learning/AI

Mark Asberg, Queen's University
Talia Chung, University of Ottawa
Scott Gillies, Wilfrid Laurier University
Vivian Lewis, McMaster University
Mark Robertson, Toronto Metropolitan University
Catherine Steeves, Western University (Chair)
Amy Greenberg, OCUL
Kate Davis, Scholars Portal
Michael Ridley, University of Guelph

Final Report – for Public Distribution

May 2024

Table of Contents

1. Introduction

The Final Report of the OCUL Task Force on Machine Learning/AI follows the Interim Report (See OCUL website) that presented a set of use cases and important contextual issues. Feedback from the community was gathered from online comments and the virtual Summit held on March 20, 2024 (The Summit Review is available on SPOTDocs – login required). Rather than incorporating the Interim Report and feedback, this report draws from those prior documents and moves directly to making specific recommendations on how OCUL can move forward with Artificial Intelligence and Machine Learning initiatives and projects.

2. Strategic Objectives

The rapidly growing field of Artificial Intelligence and Machine Learning (AI/ML) will fundamentally change the world of work, and the work of libraries. AI/ML is already being used in teaching and learning, reference, and research, and we expect this use to grow exponentially in the coming years. We strongly believe that the transformational changes offered by AI/ML will be best understood, managed, and realized if OCUL libraries work together on a select number of core initiatives, which are proposed and outlined here.

Our goal in finding, selecting, and shaping the projects outlined below is twofold. First, we aim to grow our collective expertise around specific technologies and processes through hands-on experience. The outcomes of these projects should provide a tangible benefit to the OCUL community. Secondly, we have chosen a range of projects that will help us understand scenario, workflow, and structural issues related to different types of AI/ML work. The expectation is that the *process* of doing these projects will significantly expand OCUL's ability to think through what it means to incorporate AI/ML tools into our work, and that lessons learned in these projects can be applied much more broadly in the coming years.

In short, our strategy is as follows:

1. Identify AI/ML use cases that can be implemented as operational projects for the benefit of library users, library staff or both.
2. Implement these projects as consortial, collaborative initiatives.
3. Strive to maximize the availability of tools, services, and outcomes to support both English and French speaking Ontario learners and researchers.
4. Use these projects and related work to build capacity in libraries regarding AI/ML awareness, knowledge, and technical skills.
5. Enable transformational change to further the mission of libraries.

While this document presents an OCUL strategy, a critical next step is to open initiatives and projects to partner organizations and libraries. The challenges and opportunities of AI/ML are best addressed through broad community engagement. OCUL will actively seek partnerships and foster community collaboration.

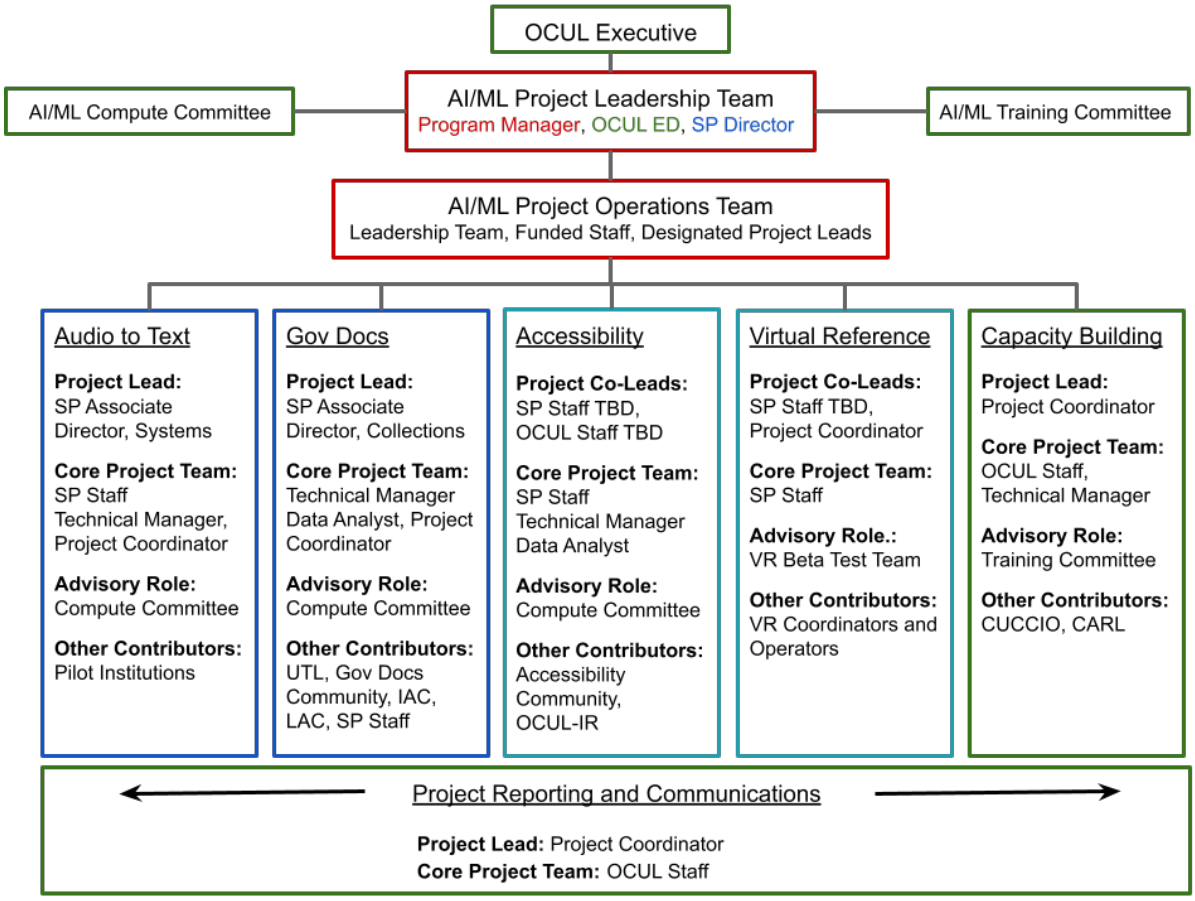
3. Guiding Principles

The OCUL Operating Principles direct that OCUL AI/ML projects and initiatives will a) encourage new ideas, enable collaboration, and build consensus, b) balance risk and opportunity, and c) be congruent with the OCUL strategic plan and its core principles. The latter highlight: advancing research, teaching, and learning; leadership in social change and inclusivity; developing and supporting robust infrastructure; and demonstrating value.

With these principles, OCUL aspires to promote "responsible AI" in the academic environment with its focus on the ethical and mindful development, deployment, and management of artificial intelligence systems that align with ethical principles, societal values, and legal regulations. Specifically, in the OCUL context, this approach:

- emphasizes collaboration,
- designs and implements with humans-in-the-loop,
- designs and builds around diversity, equity, and inclusion,
- enhances accessibility,
- advances environmental sustainability,
- ensures transparency and explainability, and
- aspires to adopt open-source tools and solutions.

4. Program Structure



Program management requires effective leadership and operations oversight. In the Governance section, see **Recommendation #3a** regarding creation of an AI/ML Project Leadership Team and an AI/ML Project Operations Team.

5. Proposed AI/ML Projects

These projects are recommended among the many use cases explored because they address several key criteria: there is an expressed need, they provide consortial solutions, they move from simpler to more complex projects allowing for learning and experience, and each project creates reusable tools and techniques for future use.

a) Audio to Text Transcription

(Lead organization: Scholars Portal)

Background & rationale: Many OCUL libraries have noteworthy collections of spoken audio files (e.g., oral histories, meetings, podcasts, speeches, broadcasts) that have not been

transcribed. Using AI/ML transcription models, these audio files can become more accessible. The resulting transcriptions are indexed to the audio for easy referencing and a corpus of related audio files could be trained for enhanced discovery. An important outcome is an alternative format that furthers AODA objectives. While transcribed audio still requires verification, one promising option is to use corrected transcripts to train (fine-tune) the model to identify and correct errors.

While there are many audio-to-text transcription tools or services available, the opportunity to explore this use case early in the Task Force investigations resulted in a pilot initiative. An open-source tool with widespread use and success (Whisper; created by OpenAI and extended by many others) was the focus of the OCUL Hackfest held on February 23, 2024

Objective: This project involves developing a Whisper pipeline at Scholars Portal, setting up the service and providing access to library staff. This will allow users to quickly begin using the tool on locally hosted servers. The project will include documentation to assist users in getting started.

Strategic learning outcome: While Whisper is of interest right now, it is likely new AI/ML tools will become available. Capacity to experiment with rapidly changing AI/ML service tools may be limited at individual schools. By centralizing and developing expertise in this process, Scholars Portal can offer OCUL libraries an opportunity to try out tools without a significant investment in time and infrastructure.

Recommendation #1a: That an audio to text transcription project be defined and initiated with a select group of libraries with diverse audio collections.

b) Government Documents

(Lead organization: Scholars Portal)

Background & rationale: Between the University of Toronto Libraries and Scholars Portal, via Internet Archive, we have a collection of about 325,000 Canadian and Ontario government documents in PDF format.

These historic documents, some as early as the 1840s, have relevance to many research fields. Unfortunately, searchable metadata is often either unavailable or uneven in extent and quality. Providing enhanced metadata and new discovery tools would substantially increase accessibility and useability of these resources.

Objective: Work with community partners (LAC, IAC, OCUL Government Information Community) to improve metadata for much of the government documents collection, including more accurate and cleaner titles, series, and producing bodies. Adding summaries (with learning from the ACE project) and developing links between different government bodies.

Strategic learning outcome: Government documents are a rich and valuable resource and experience creating and enriching metadata using ML can be applied to many other

kinds of scans or archival collections such as Canadiana. We will gain deep experience in scoping fields, and in training and refining discovery tools.

Recommendation #1b: That a government documents project be defined and implemented to enhance metadata, discovery, and analysis with a broad, if selective, use of corpora available to OCUL.

c) Accessibility: Accessible Content E-Portal (ACE) Book Summaries and Document Remediation

(Lead organizations: OCUL Office/Scholars Portal)

Background & rationale: The Accessible Content E-Portal (ACE) makes available scanned print books in multiple formats to users at universities and colleges in Ontario with print disabilities. Users have said it can be difficult and time consuming to determine whether particular titles meet their needs without downloading the whole file. They are interested in book summaries. These could be done at the book and/or chapter levels.

Additionally, the quality of the scans often means that additional remediation is needed to meet user needs (e.g., to make the documents accessible to screen readers). There are an ever-growing number of AI tools available to help with document remediation, some are available through licensing while others could be hosted locally. These options should be explored and will determine the expertise needed for the project.

Objective: Increase usability of the ACE portal by generating and adding summaries (ideally at the chapter level) to scanned book metadata. Select and begin the process of working with an AI/ML remediation tool.

Strategic learning outcome: Summaries may be good in many other contexts, and remediation may be helpful for all kinds of archival or messy content, so both parts of this project could be expanded upon. We will also gain an understanding of any licensing implications of doing this kind of enhancement or remediation. While in the ACE-specific context this is less of a concern, this may be a good springboard for exploring these issues with other collections.

Recommendation #1c: That two ACE subprojects be defined and implemented: 1) an ACE book summaries project using open-source tools to create abstractive summaries and add them to the item metadata, and 2) an ACE document remediation project prioritizing research into human-in-the-loop techniques and options that could be adopted or adapted to ACE requirements.

d) Virtual Reference

(Lead organizations: OCUL Office/Scholars Portal)

Background & rationale: Response to the Virtual Reference (VR) use case in the Interim Report noted both opportunities and challenges. While some saw a new service with extended reach (and something that might deflect students away from generic chat services), others had concerns about quality and human agency. A pedagogical approach to VR with a strong "human-in-the-loop" strategy was widely supported.

LibraryH3lp, OCUL's Ask a Librarian platform, is in the early stages of developing a separate VR chat product initially designed as a tool for "operators" (i.e., staff providing the Ask a Librarian service) and eventually to be released as a standalone service. OCUL has been asked to provide Ask a Librarian logs for training and to work with LibraryH3lp as beta testers of their VR product.

LibraryH3lp is a small company with modest AI/ML experience. However, working with them offers OCUL a low-risk opportunity with a limited resource commitment. OCUL could easily pivot to another strategy (e.g., a different provider or local development) should the LibraryH3lp product be deemed unsatisfactory.

Objective: A chatbot, owned by LibraryH3lp, that has experience with university website data. The objective here is not to begin using a chatbot, but to participate in its development and then perform an analysis of its ability. This tool must be human-mediated, give consistently high-quality responses, and have some type of information literacy objectives built into its process.

Strategic learning outcome: Understanding the validity of chatbot tools, how well they can be trained, and in which contexts they may be appropriate. This will be a useful case study in working with a vendor to develop an AI tool. More broadly, OCUL should use this project to better understand and meaningfully incorporate concerns into project planning and development related to AI and reference.

Recommendation #1d: That OCUL work with LibraryH3lp to assist in development of their Virtual Reference chat product by sharing publicly available data and acting as beta testers.

In the Governance section, see **Recommendation #3b** regarding creation of an OCUL Virtual Reference Beta Test Team.

e) Capacity Building Across OCUL

(Lead organization: OCUL Office)

Background & rationale: Positioning capacity building as one of the OCUL AI/ML projects underscores its significance to the immediate and ongoing success of the overall strategy.

The recommended projects are designed to build capacity as they progress from relatively simple to more complex. This includes compute capacity, but it also involves increasing technical expertise, deepening an understanding of the challenges of reimaging service and resource provision with AI/ML, and furthering responses to key contextual issues such as copyright, bias, and veracity.

However, adopting, providing, supporting, and developing AI/ML tools, resources, and services as envisioned in this strategy requires a broader and focused effort on capacity building.

Objectives: Enhancing staff competencies, as users, providers, and developers of this technology and as teaching and learning professionals, is designed around five components:

e.1) *Staff AI/ML Training*: AI literacy and skill training for all library staff to support their current and future work.
e.2) *Advanced AI/ML Technical Training*: AI/ML training for technically focused staff to contribute to OCUL and local projects.
e.3) *AI Literacy and Instruction*: facilitating AI literacy curriculum and learning strategies for students and faculty in concert with local teaching and learning initiatives.
e.4) *AI/ML Tools and Service Licensing*: enhancing expertise in licensing and acquiring AI/ML tools and services to facilitate the review and acquisition of emerging products.
e.5) *Engaging Students*: in consultation with the LIS Deans and faculty engage undergraduate and graduate students in OCUL AI/ML initiatives.

Strategic learning outcome: Staff at OCUL libraries have a high level of AI/ML understanding and experience. Their expertise allows them to use, develop, and provide instruction regarding AI/ML as appropriate for their roles.

e.1) Staff AI/ML Training

While there is general agreement with the need to move OCUL libraries from a passive to an active position with respect to AI and machine learning, a common concern is the need for staff training to enable this. Training requirements ranged from general awareness and AI literacy to the deep technical knowledge and expertise needed to design and implement applications.

Library staff already encounter AI/ML tools, services, and resources in their daily work. AI/ML training will assist them with the current challenges and opportunities and prepare them for the future as AI/ML continues to evolve.

Many resources exist for awareness, AI literacy, and general technical proficiency. The need is less creating learning programs than identifying existing programs and making them available to the OCUL library workforce. Programs such as DestinationAI, MinnaLearn and the University of Helsinki's Elements of AI, and those offered from the Library Carpentry are relevant examples.

Collaborative learning is often the most effective. The OCUL Whisper Hackfest was very successful in offering a learning and development environment suitable for staff of all backgrounds and skill levels. Additional hackfests (whether in-person or online) could be offered to address other applications, key issues, or emerging services.

To enable staff AI/ML training at these different levels and for these different requirements, suggests the need for a group within OCUL to coordinate training resources and opportunities, and to liaise with groups that have similar objectives.

In the Governance section, see **Recommendation #3c** regarding creation of an OCUL AI/ML Training Coordination Committee.

e.2) Advanced AI/ML Technical Training

Currently few libraries in OCUL have technically focused staff with much experience with AI/ML. Developing more comprehensive technical expertise in AI/ML is important if libraries are to maximize the value of this technology.

Providing deeper technical training in AI/ML for staff is a challenge shared with university central IT departments in Ontario and across the country. Coordinating technical training opportunities with Canadian CIOs through the Canadian University Council of Chief Information Officers (CUCCIO) and through organizations like Compute Ontario offers not only a cost effective opportunity but also serves to highlight library issues to these groups. Preliminary discussions with Lindsay Sill, CUCCIO Executive Director, indicate a strong interest in collaborating on this initiative. Engaging CARL in this proposal would facilitate this as a national program.

Recommendation #1e: That the OCUL AI/ML Program Manager and the Project Coordinator initiate discussions with CUCCIO and CARL to explore the feasibility and options for a national initiative for AI/ML training for technically focused staff in academic libraries and central IT organizations.

e.3) AI Literacy and Instruction

The need for AI literacy and instruction for students and faculty was widely identified during the consultation phase. Many librarians are working closely with their institutional teaching and learning groups and are involved in committees to assess the impacts and opportunities of AI/ML. The work of the recently created subcommittee on AI in teaching and learning by the CARL Advancing Teaching & Learning Committee will be important to these efforts. These engagements are also opportunities for consideration of the larger issues of information literacy, trust, misinformation, and attribution.

Given the specific, and often unique, cultures and policies regarding teaching and learning in universities, the role of OCUL should be in supporting information sharing and facilitating collaboration among peers.

AIMLComm, the community of practice regarding AI/ML in academic libraries, has a channel focused on AI literacy and instruction with over 100 participants. Register online to join the *AIMLComm*. In addition to information sharing, discussions on this channel could propose a hackfest relevant to curriculum development or to initiate regular peer meetings (Zoom/Teams) to explore experiences, strategies, and challenges.

Recommendation #1f: That the *AI Literacy and Instruction* channel in the *AIMLComm* community of practice be promoted by OCUL as a place for information sharing and collaboration among the library workforce engaged in teaching and learning initiatives regarding AI/ML.

e.4) AI/ML Tool and Service Licensing

AI/ML is being incorporated into many information resource products, some existing and others new. Examples include scite.ai (smart citations and literature search), Scopus AI (scholarly resource database), Transkribus (manuscript transcription), and DistillerSR (systematic reviewing). Another interesting example is Constellate from ITHAKA, an aggregation of tools and code to perform various text analysis functions.

Evaluating and acquiring or licensing these tools and resources will be of interest to many OCUL libraries and there may be opportunities for consortial acquisition.

It is recommended that such decisions follow the existing process enabled by the OCUL Information Resources Committee (OCUL-IR). Positioning OCUL-IR in this new role would facilitate the process of collecting interest, establishing evaluation criteria, determining benefits and acquisition details, and moving forward with negotiations by OCUL staff should the situation warrant it. Given that these tools differ from resources typically reviewed by OCUL-IR, it is suggested that those seeking consortial licenses work with OCUL-IR to help define the need and criteria. In consultation with the OCUL Executive, OCUL-IR should consider updating its mandate and composition to include a focus on tools and services, and to add new members or a new subcommittee to address the specific challenges of licensing tools and resources.

Recommendation #1g: That the OCUL Information Resources Committee (OCUL-IR) take on the responsibility for licensing or acquiring AI/ML tools or services by collecting interest, establishing evaluation criteria, determining benefits and acquisition details, and moving recommendations forward to OCUL staff to initiate negotiations should the situation warrant.

e.5) Engaging Students

Students from library and information science programs with relevant AI/ML or related experience could be involved with specific projects for a limited duration. These arrangements could be internships, practicums or short term, part-time hires.

Engaging students not only provides them with a valuable experiential opportunity, but it also offers OCUL libraries a pipeline to potential recruits with highly relevant technical expertise.

Assessing expertise, coordinating recruitment, managing work, time constraints, and arranging funding and/or course credit would be some of the considerations. However, giving students real world work experience would benefit all parties. While initial contacts with LIS school deans and faculty have explored this option, additional discussions and planning are required. Broadening student interest from other programs (e.g. computer science) could be a second phase of this initiative.

Recommendation #1h: That the OCUL AI/ML Program Manager pursue options with LIS school deans and faculty to engage undergraduate and graduate students in OCUL initiatives.

6. Resourcing Projects

a) People

In order to advance the work proposed in this project, OCUL will need to 1) reallocate existing OCUL office and Scholars Portal staff, 2) add new capacity to the OCUL office and Scholars Portal teams, and 3) engage staff from OCUL member libraries.

a.1) In-kind Contributions

Program Manager: The overall OCUL AI/ML strategy requires oversight to ensure appropriate guidance, direction, liaison (with OCUL, Scholars Portal, and other partners), and assessment. Such a position requires an understanding of AI/ML in the academic library context and senior administrative experience to manage a complex strategy. Catherine Steeves (Western University) has been appointed to this role.

OCUL and Scholars Portal Directors: Involvement of the Directors of both OCUL and Scholars Portal is required to ensure alignment and integration with current programs and will help build the foundation for future work in both organizations.

OCUL Office and Scholars Portal Team Members: Every project that has been identified will require contributions from current OCUL Office or Scholars Portal staff members. OCUL Office and Scholars Portal staff will provide administrative and communications support, expertise in community engagement, licensing, shared infrastructure provision, metadata, and service and platform development.

OCUL Member Libraries Staff: Staff from OCUL member libraries will play a vital role in the success of these projects via work on committees and working groups, participation on project teams, and piloting of services. For many libraries this will require challenging conversations and decisions about library priorities and staff reallocations.

a.2) Funded Positions

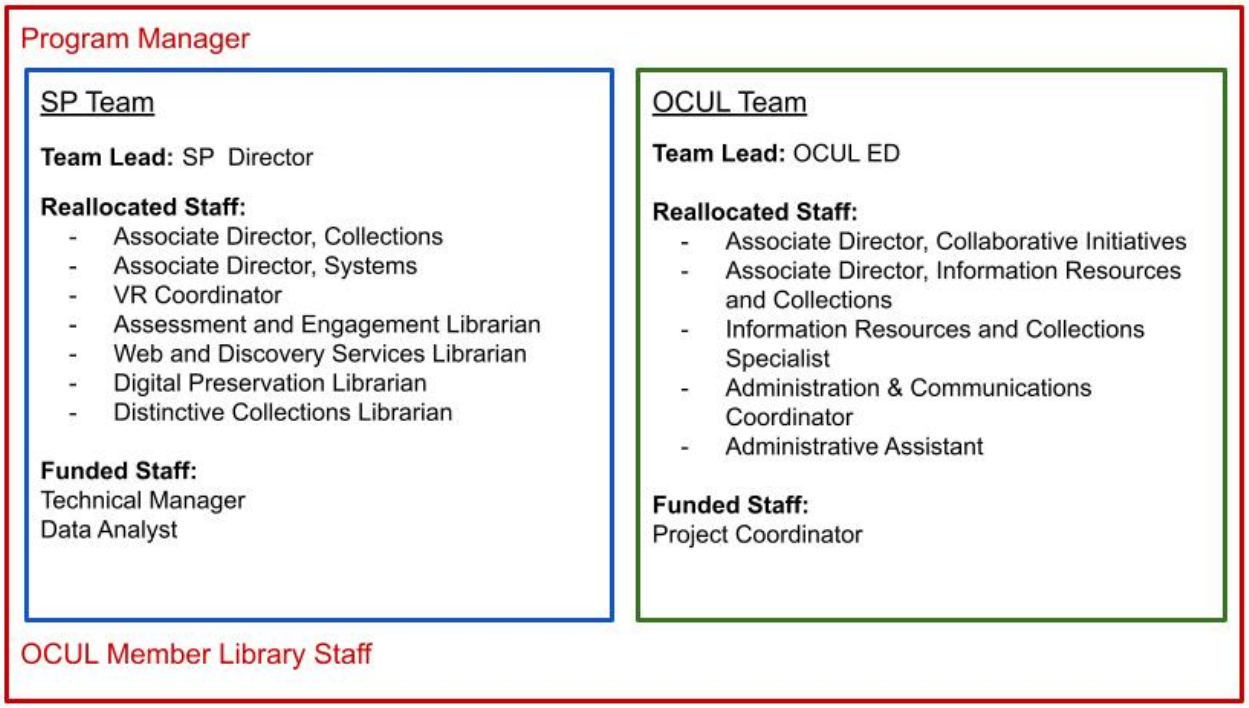
Project Coordinator: As with all large undertakings with multiple goals, diverse project teams, and complex resource allocations, the OCUL AI/ML initiative requires a Project Coordinator to facilitate success. This role will be involved in initiating, planning, monitoring, and ensuring quality control across AI/ML projects and initiatives. They will support the work of any AI/ML committees or working groups, coordinate with project teams and community members, communicate project activities and milestones within OCUL, and lead the operations of the Capacity Building project. They will also be instrumental in establishing sustainability plans for current projects, and in assessing and evaluating the work of project teams.

Technical Manager: This role will take the lead in identifying and using appropriate AI/ML techniques as part of the proposed projects. They will lead technical aspects of projects and advise on effective ways to build necessary infrastructure. They will take the lead on building partnerships that help select and integrate tools and services, with a focus on working with national and local compute infrastructure. They will also be instrumental in planning and delivering training for technical staff.

Data Analyst: Both the Government Documents project and the Accessibility project will require working with large quantities of data. The ideal candidate will have experience with AI/ML tools and platforms, skills in Python coding and data analysis, and will advise on related projects as needed.

a.3) Summary of Core Project Team

The core project team will consist of the Program Manager, SP and OCUL staff contributing part of their time, several full-time funded positions, and committees and working groups of OCUL member library staff.



b) Compute Resources

There are a number of options to access the compute resources necessary to implement the OCUL AI/ML strategies. Both the Digital Research Alliance of Canada and Compute Ontario offer services and resources that are freely available to Canadian researchers. Access to GPU compute is available as needed (up to a finite threshold). It is anticipated that this compute limit will not be problematic for the proposed projects.

Another option is to engage with a full-service commercial provider, such as Amazon's AWS or Microsoft's Azure, and utilize their standard processes (including access to relevant software and compute requirements). This approach would allow OCUL to draw on the expertise available through the support services from these companies. However, the cost models for these services are complex since each component and process is costed separately. AWS and Azure offer some "free tier" or academic friendly services that could be a cost-effective option for small-scale pilots.

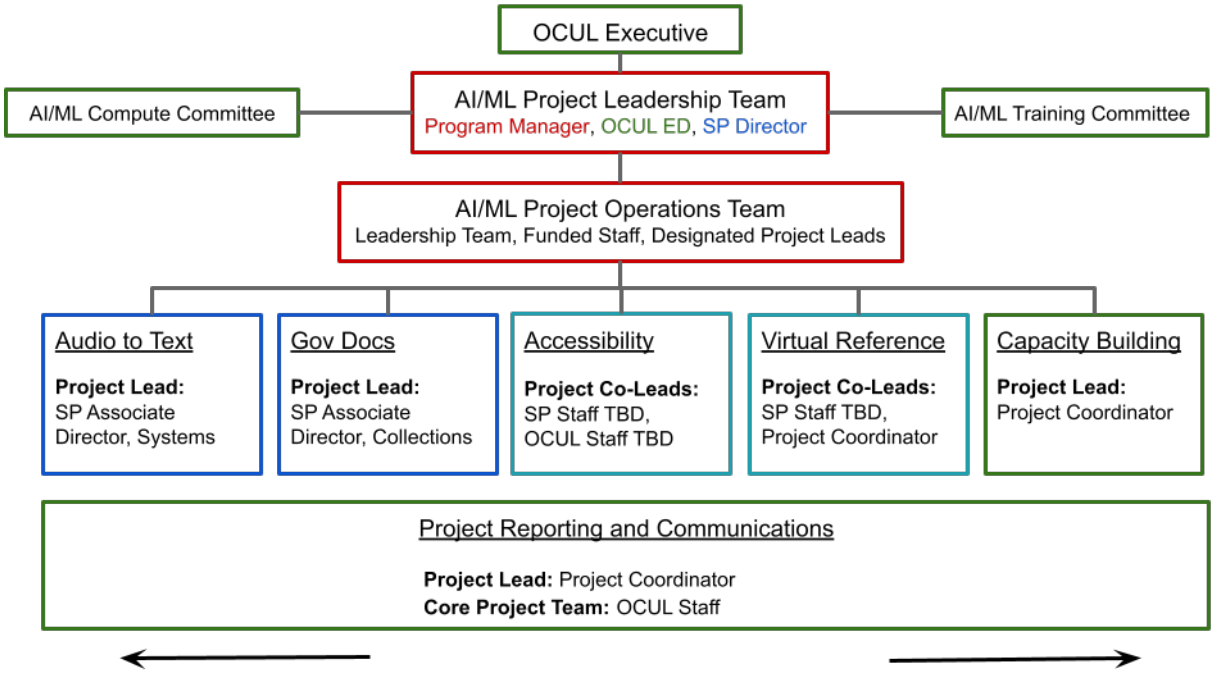
Scholars Portal has existing compute resources that can be made available for specific and selective use. It may be desirable to invest in further capacity in order to enable more flexible local processing of data.

The effective tracking, assessing, and advising on compute resources and associated tools and services necessary to support OCUL AI/ML projects requires an oversight committee to engage in this work.

In the Governance section, see **Recommendation #3d** regarding creation of an OCUL AI/ML Compute Committee.

c) Governance

To ensure the identified projects are successful and meet the needs of OCUL libraries, appropriate governance needs to be put in place.



Recommendation #3a: That an **AI/ML Project Leadership Team** be defined as Program Manager, OCUL Executive Director, and Scholars Portal Director and that this leadership team establish an **AI/ML Project Operations Team** to coordinate overall project activities. The Operations Team would include the Leadership Team, Project Staff (Technical Manager, Project Coordinator), designated Project Leads and would include additional staff resources as appropriate. The **AI/ML Project Leadership Team** will report to the OCUL Executive.

Recommendation #3b: That an **OCUL Virtual Reference Beta Test Team** be established drawn from members of the Virtual Reference Steering Committee and from OCUL library staff with relevant interest and expertise. The **OCUL Virtual Reference Beta Test Team** would report to the **AI/ML Project Operations Team**, with the OCUL Virtual Reference Coordinators Committee playing an important consultative role.

Recommendation #3c: That OCUL mandate and launch an **OCUL AI/ML Training Coordination Committee** who will work with the Project Coordinator to identify and secure AI/ML awareness and training opportunities relevant to the OCUL library workforce. The OCUL AI/ML Training Coordination Committee would report to AI/ML Project Leadership Team and, where possible, include representation from other OCUL committees (Information Resources, Scholars Portal, and Collaborative Futures).

Recommendation #3d: That OCUL mandate and launch an **OCUL AI/ML Compute Committee** to track, assess, and advise regarding alternative resources, models, frameworks, compute, and associated tools. The OCUL AI/ML Compute Committee would report to the AI/ML Project Leadership Team and, where possible, include representation from other OCUL committees (Information Resources, Scholars Portal, and Collaborative Futures).

d) Funding

OCUL directors approved allocations from the OCUL New Initiatives Fund, providing funding over 2024-2025 and 2025-2026 for start-up costs.

In addition to this funding, in-kind contributions from OCUL libraries will form a crucial part of the funding strategy. This includes staff serving on committees and working groups to provide both technical and domain expertise.

Engaging with partners outside OCUL, especially for specific project use cases, is anticipated. These partnerships might involve shared costs and in-kind contributions, with potential partners including commercial groups as well as research libraries and related organizations. Formal co-development agreements will be essential.

Opportunities exist to obtain funding for specific AI/ML initiatives, particularly through SSHRC's Insight Development Grants and Partnership Engage Grants. Another possible funding source, for AI literacy and instruction projects, is eCampusOntario. The specifics of these funding opportunities warrant further exploration by the Program Manager.

7. Timeline

The following timeline is selective and illustrative but reflects the specific recommendation that work begin as soon as approval is obtained.

May 2024
Approval to proceed from OCUL Directors (Spring Meeting)

June-August 2024
The project leadership team:

- Names/hires OCUL AI/ML management and technical staff
- Mandates and establishes OCUL Virtual Reference Beta Test Team
- Mandates and establishes OCUL AI/ML Compute Committee
- Mandates and establishes OCUL AI/ML Training Coordination Committee

September 2024
Scholars Portal team, with appropriate consultation:

- Defines and initiates the Audio to Text Transcription Project
- Defines and initiates the Government Documents Project

October-November 2024
Project leadership team submits a status report to the OCUL Directors (Fall Meeting)

November 2024
OCUL-IR mandate reviewed and capacity for AI/ML licensing initiated
ACE Accessibility Projects defined and initiated

December-March 2025
Complete Audio to Text Transcription Project
Project work continues

April-May 2025
Project leadership team submits a year-end report to the OCUL Directors (Spring Meeting)

May-August 2025
Project work continues

September-December 2025
Project work continues
Consideration of sustainability plan and possible future projects is initiated
Project leadership team submits a status report to the OCUL Directors (Fall Meeting)

January-April 2026
Project work wraps up
Project leadership team submits final report to the OCUL Directors (Spring Meeting)